

NCC2-387

7N-82-CR

280697

PM

## Electronic Publishing

*Peter J. Denning*

October 10, 1986

Research Institute for Advanced Computer Science  
NASA Ames Research Center

RIACS Technical Report 86.21

(NASA-CR-180683) ELECTRONIC PUBLISHING  
(Research Inst. for Advanced Computer  
Science) 19 p

N91-70395

00/82 Unclass  
0280697

# RIACS

Research Institute for Advanced Computer Science



## Electronic Publishing

*Peter J. Denning*

Research Institute for Advanced Computer Science  
NASA Ames Research Center

RIACS Technical Report 86.21  
October 10, 1986

**The technology for desktop publishing is producing new opportunities for printing and distributing scientific documents. By the same token, it threatens scientists with an ever-increasing flow of mediocre publications. This article is a summary of progress in six crucial areas: manuscript preparation, printing, electronic submission, electronic distribution, integration with the research environment, and aesthetics.**

---

Work reported herein was supported in part by Cooperative Agreement NCC 2-387  
between the National Aeronautics and Space Administration (NASA)  
and the Universities Space Research Association (USRA).

This is a preprint of the column *The Science of Computing* for  
*American Scientist* 74, No 6, November-December, 1986.

---

ORIGINAL PAGE IS  
OF POOR QUALITY



## Electronic Publishing

Peter J. Denning

Research Institute for Advanced Computer Science

October 10, 1986

"Desktop publishing house!" A personal computer and laser printer, together with software for word processing, graphs, charts, database, and spreadsheet, so say the ads, can enable you to create camera-ready documents of professional quality in your business or home. A small personal computer system with a slow laser printer might cost \$10,000, while a workstation and a fast laser printer might cost \$30,000. How useful are these systems for scientific publishing?

To explore this question, I will consider technical publication in its broad sense: the process leading from the gathering of information during research and development to the appearance of that information in a magazine or journal. For a published paper, this process usually consists of the following eight steps involving author, editor, and publisher:



1. Author performs experiments, analyzes data, generates findings.
2. Author prepares manuscript, including text and graphs, tables, and other supporting materials.
3. Author sends manuscript to editor.
4. Editor obtains reviews, informs author of recommendations.
5. Author incorporates revisions, returns manuscript to editor.
6. Editor forwards accepted paper to publisher.
7. Publisher edits and formats manuscript, sets galleys, obtains author's approval of galleys.
8. Publisher makes press run, distributes copies.

Typically four computers are involved in the process: for performing the scientific work (Step 1), for preparing and revising the manuscript (Steps 2 and 5), for preparing a copyedited version of the manuscript (Step 7), and for printing (Step 8). These computers are likely to be independent and incompatible. Until the computers conform to common standards and are linked together in networks, full electronic publication of scientific material will remain a distant goal.

Some will argue that standards and networks are not the whole answer; changes in the publication process itself may be needed, especially changes that eliminate editing and publishing middlemen. Changes of this kind have already begun in networked communities with the institution of unedited, unreviewed electronic bulletin boards on specific subjects. Yet, wholesale replacement of edi-

tors and publishers appears unlikely. As long as we use a system of peer review to evaluate manuscripts, editors will be an integral part of the process. As long as printed paper remains cheaper, more portable, and more usable than electronic media, people will prefer it, and publishers will be an integral part of the process because they can provide high print quality and handle bulk distributions at the lowest cost. On the other hand, changes in the ways editors and publishers perform their tasks are likely. I will discuss some possibilities below.

Where are we now on the road to electronic publication? What follows is a summary of progress in six crucial areas: manuscript preparation, printing, electronic submission, electronic distribution, integration with the research environment, and aesthetics.

Systems for manuscript preparation have become quite sophisticated. When coupled with high-resolution laser printers, the best do indeed produce results of professional caliber. When coupled with low-cost, medium-resolution laser printers, they provide a creditable desktop printing facility.

The traditional approach to manuscript preparation is at least as old as a program called RUNOFF, developed at MIT in the early 1960s. The idea is to embed format commands within a document's source file, which is prepared with any simple text editor; the source file can be given as input to a formatter program that generates a file of printer commands; the printer program takes that file and a file of fonts, and prints the document. The formatter program takes care of margins, selection of fonts, centering, indenting, page numbers,

paragraphing, filling and justification of lines, hyphenation, and footnotes.

I will use UNIX, the system most familiar to me, as a source of examples. In UNIX, documents can be prepared with the help of a series of six programs, each of which has its own specialized language. The formatter, *troff* (usually pronounced "tee-roff"), is a descendant of RUNOFF. The other programs -- *tbl*, *eqn*, *pic*, *grap*, and *refer* -- are preprocessors for *troff*. They look for marked portions of the source file containing statements in languages that specify tables, equations, pictures, graphs, or citations; they replace these portions with formatter commands. Samples of input and output for text, tables, and equations are shown in the first three figures.

To use these programs, an author must become proficient in the six specialized languages mentioned above. Relatively few authors accomplish this. Moreover, many authors wind up expending considerable effort to debug their documents -- that is, to fine tune formatter commands to improve a document's appearance. I refer to the traditional style of manuscript preparation as the programmer's approach.

Two other widely used traditional programs are Scribe, developed by Brian Reid in 1978, and T<sub>E</sub>X (pronounced "tech"), developed by Donald Knuth in the same year (1). Scribe has a much simpler and more uniform syntax than *troff*, a comparable table facility, and weaker graphics and math subsystems. T<sub>E</sub>X has a more powerful math subsystem and no graphics; it uses a new font called computer modern and produces highly pleasing results. T<sub>E</sub>X is used by the



## FIGURE 1

```
.ce 2
.sz 14
\fbElectronic Manuscript Preparation\fr
\fiPeter J. Denning\fr
.sz 12
.sp 2
.pp
This issue of \fiAmerican Scientist\fr contains an essay about
electronic publishing. The essay contains samples of source files
prepared under the UNIX\us-4TM\s0\d \fitroff\fr program, together
with their corresponding outputs.
.pp
The samples illustrate paragraphing, different point sizes, bold and
italic fonts, automatic line filling, and right adjustment.\(dg
.(f
\((dgThey also illustrate footnotes, which are automatically put at
the bottom of the page in a smaller point size.
.)f
The source file contains a large number of format commands. Some are
single lines beginning with dot; others are embedded in the text
preceded by backslash (\). As a result, the source file is
difficult to read.
.pp
This particular version was printed on a 300 dpi QMS laser printer
using a font called "Tolares Roman."
```

### Electronic Manuscript Preparation *Peter J. Denning*

This issue of *American Scientist* contains an essay about electronic publishing. The essay contains samples of source files prepared under the UNIX<sup>TM</sup> *troff* program, together with their corresponding outputs.

The samples illustrate paragraphing, different point sizes, bold and italic fonts, automatic line filling, and right adjustment.<sup>†</sup> The source file contains a large number of format commands. Some are single lines beginning with dot; others are embedded in the text preceded by backslash (\). As a result, the source file is difficult to read.

This particular version was printed on a 300 dpi QMS laser printer using a font called "Tolares Roman."

---

<sup>†</sup>They also illustrate footnotes, which are automatically put at the bottom of the page in a smaller point size.

FIGURE 2

```
.EQ
gsize 12
delim @@
EN
EQ (1)
{\partial \sup 2 u} over {\partial t \sup 2} ~~~~~
sum from j=0 to n t
{\partial \sup 2 u} over {\partial x sub j \sup 2}
EN
.sp
where
EQ (2)
u ~~~~~
sum from {roman all ~ bold n} ~
prod from i=1 to K ~
V sub i (N) S sub i (n sub i ) !
EN
.sp
with @bold n = (n sub 1 , ..., n sub K )@
and @n sub 1 + ... + n sub K = N@.
```

$$\frac{\partial^2 u}{\partial t^2} = \sum_{j=0}^n \frac{\partial^2 u}{\partial x_j^2} \quad (1)$$

where

$$u = \sum_{\text{all } \mathbf{n}} \prod_{i=1}^K V_i(N) S_i(n_i)! \quad (2)$$

with  $\mathbf{n}=(n_1, \dots, n_K)$  and  $n_1 + \dots + n_K = N$ .

FIGURE 3

```
.TS
doublebox center tab(:);
cp14 s , cp14 s , lB | cB , l | c .
.sp .2v
DIRECTORY
SCHOOL OF OLOGY

Department:Room

.sp .2v
Anthrop:301
Archae:126
Bacteri:109
Bi:326
Entom:217
Etym:221
Ge:204
Paleont:113
Physi:312
Psych:209
Toxic:307
Zo:117
.sp .2v
.TE
.ce 2
.sz 10
Source: S. Harris, \fIAmerican Scientist\fR,
July-August 1986, p. 365. liked a 300 dpi printer.
```

DIRECTORY SCHOOL OF OLOGY	
Department	Room
Anthrop	301
Archae	126
Bacteri	109
Bi	326
Entom	217
Etym	221
Ge	204
Paleont	113
Physi	312
Psych	209
Toxic	307
Zo	117

Source: S. Harris, *American Scientist*,  
July-August 1986, p. 365. liked a 300 dpi printer.

American Mathematical Society for publication of its journals.

The difficulties of using the programmer's approach have inspired another, which I call the writer's approach. Most personal computers use this second approach, and people who have never worked with such programs as *troff*, Scribe, or  $\text{T}_{\text{E}}\text{X}$  are often surprised to learn that there is any approach other than the writer's. Computer people call the writer's approach WYSIWYG (pronounced "wizzy-wig"), an acronym for "what you see is what you get." It originated with the Bravo editor at the Xerox Palo Alto Research Center in the mid-1970s. The idea is that the author sees nothing but the formatter's output, constantly updated in real time, on the display screen. Editing requests are applied directly to the image of the document; in response, the system inserts text and formatting codes in a file never seen by the author. A print command causes the printer to produce exactly the image shown on the display.

At present, many WYSIWYG editor programs suffer from annoying limitations, such as inability to handle large files, to incorporate pictures, charts, tables, and equations into documents, or to take data for tables, charts, and graphs from separate files. They generally do not offer the same degree of control over the quality of tables, charts, and equations as do traditional formatters, but they are improving steadily.

Printing, the second crucial area, involves the use of a formatter's output to drive a devices ranging range from "letter quality" dot matrix printers to laser printers and phototypesetters. The patterns of small dots impressed on the

paper by a dot matrix printer are too coarse to produce aesthetically pleasing documents. The more sophisticated laser printer consists of a microprocessor, memory for holding fonts and the data to be printed, and a "marking engine"; the marking engine uses a laser to imprint images on paper, the laser's control signals coming from the microprocessor rather than from an optical scanner. Laser printers come in many grades. The least expensive ones cost around \$2000, have resolutions around 300 dots per inch (dpi), and are capable of printing one text page, without graphics, in about 9 seconds; slow links to slow personal computers, however, are likely to increase page-print times to minutes, especially if pictures appear. A 300-dpi printer capable of producing one text page every 5 seconds is likely to cost at least \$15,000. Professional phototypesetters, such as the ones used to produce the camera-ready copy for *American Scientist*, have resolutions on the order of 1200 dpi. The cheapest cost around \$35,000, and 2540-dpi models are twice that. A high-resolution printer requires high-resolution fonts; with a low-resolution font, it will perform no better than a 300 dpi printer.

The manufacturers of printers have been interested in standardizing the languages in which printer input (formatter output) is expressed. The example most common in the United States is PostScript. Another, called ACE, is used by at least one publisher in England to distribute the contents of its newspaper to regional printing centers.

Electronic submission, the third area, refers to the transfer of a manuscript electronically from author to editor, from editor to reviewer, or from editor to publisher. The file must be stored in a standard text code (e.g., ASCII) that can be read by different computers; it can be transferred on a floppy disk, over a telephone connection, or over a network. Some WYSIWYG editors store format information as non-ASCII codes in the source file, in many cases preventing their transmission over phone connections and networks. Nevertheless, within a few years, all format codes will be network compatible, and networks will be an important medium of manuscript transfer.

After the networks are in place, the next obstacle is a lack of standards for describing documents and their components. How are paragraphs, sections, fonts, tables, charts, graphs, citations, and the like to be represented? How many markup languages -- the name for description languages in the publishing trade -- should publishers support and editors allow? Languages for describing documents are called markup languages in the publishing profession. Formatter input languages, such as *troff*, Scribe, and T<sub>E</sub>X, are examples. So are the files used by WYSIWYG editors. IBM uses one called Standard Generalized Markup Language. At present, the different kinds of markup languages are not interchangeable. A summary of efforts to develop international standards has been given by Wolfgang Horak (2).

Even today, in the absence of standards, many publishers are pleased to receive a copy of a source file; all they need to do is manually strip out the for-

mat commands and insert their own. Often called "capturing the author's keystrokes," this expedient saves work and removes the possibility of errors that normally occur when the publisher retypes the manuscript.

There is another side to this question. Even supposing that manuscripts are submitted in standard formats, will purely electronic forms be accepted by editors and reviewers? Editors and reviewers say that manuscript evaluation is most efficient when they have printed copies. Is it realistic to expect them to make their own paper copies, or to work only with displayed images?

While it is difficult to envisage significant departures from paper copy during peer review, it is easy to imagine that computers could support the review process in other ways. Editors could obtain names of reviewers from a common database, thereby spreading the reviewing load more evenly and achieving better matches between papers and reviewers. An editor could query proposed reviewers by electronic mail to determine whether they were ready and willing to evaluate the manuscripts in question. Reviewers could return their reports by electronic mail.

Electronic distribution -- our fourth topic -- means the dissemination of the contents of journals by electronic means rather than on printed paper. No refereed journal or commercial magazine is now distributed electronically.

In what ways might electronic distribution be practical? A 100-page issue of *American Scientist*, printed at a resolution of 1200 dpi, contains on the order of  $10^{12}$  bits. A facsimile with a resolution of 300 dpi (i.e., about  $10^8$  bits), sent

over a 1200 baud telephone link with 10-to-1 image compression, would take about 2.3 days of continuous transmission, amassing a long-distance phone bill of about \$830. If the magazine were described in a markup language, it could be transmitted in around 2 hours (phone bill about \$30), and could be converted to printed copy by a receiving computer within a few hours. Who would pay these connect charges or hardware costs, when a printed copy can arrive in the mail for \$1?

A more practical approach would be based on electronic queries by subscribers. The subscribers would, at their convenience, connect to the publisher's computer and examine tables of contents and abstracts; they could then order either printed or electronic copies of interesting articles. Whether such an enterprise could survive in the absence of advertising -- a major source of support for many magazines -- is an open question.

David Gifford of the MIT Laboratory for Computer Science has designed a more radical scheme. Information is broadcast over a radio station and can be received by personal computers with special decoders. This scheme is most likely to be used where the information appeals to a wide segment of the population -- for example, news services. I expect that other significant pilot experiments in on-line journalism will begin in a few years as the community of scientists connected to networks enlarges.

In the area of integration of publishing and research environments, we find that what is required is the capability of switching and transferring information

easily between the modes of performing research and preparing manuscripts. At present, these two modes are usually assigned to different computers. A scientist might run research programs on a supercomputer and then manually insert portions of printouts and plots into technical reports on a minicomputer. With the ability to place the results of research computations in files that can be input to manuscript preparation programs, data for a graph or a table can be incorporated automatically into a document at the proper place.

What about information flow in the other direction -- from manuscript preparation to research computing? One possibility is illustrated by Donald Knuth's WEB system, in operation at Stanford University since 1983 (1,9). This system allows researchers to generate scientific programs in languages that operate at a much higher level of abstraction than current programming languages like Pascal or Fortran. A researcher can write a technical report about a problem and the computational methods for its solution; the report contains code segments that are later extracted by a special compiler and formed into a Pascal or Fortran program. A similar concept is being pursued in the Gibbs project at Cornell University, under the direction of Ken Wilson.

Our sixth and final area of concern is aesthetics. This a side to the publishing process, which, sadly, is often neglected or played down by technical people.

The personal-computer market is flooded with fonts that are irritating and even difficult to read (see the fourth figure). Providing little rhythm or flow, and awkwardly mixing black and white space, they appear to have been designed



FIGURE 4

## Electronic Manuscript Preparation

*Peter J. Denning*

This issue of *American Scientist* contains an essay about electronic publishing. The essay contains samples of source files prepared under the UNIX<sup>TM</sup> *troff* program, together with their corresponding outputs.

The samples illustrates paragraphing, different point sizes, bold and italic fonts, automatic line filling, and right adjustment.\* This particular version was prepared with Microsoft Word<sup>TM</sup>, a WYSIWYG system, on an Apple Macintosh<sup>TM</sup>. The source file is invisible. The document was printed as displayed on an Apple Imagewriter<sup>TM</sup>, a dot matrix pinter, using a font designed by Apple called "New York." The author believes this font is difficult to read.

---

\*They also illustrate footnotes, which are automatically put at the bottom of the page in a smaller point size.

ORIGINAL PAGE IS  
OF POOR QUALITY

in haste, without benefit of professional designers. Recognizing that formatting and font design are inseparable, Donald Knuth developed Metafont, a software system for font design (1). An excellent summary of recent attempts to create fonts that will work on display terminals and laser printers of different resolutions has been given by Carolyn Chauncey (4). Some of the better fonts are on the market now, but you have to hunt for them. I expect the situation to improve only gradually.

A second area in which aesthetics is succumbing to amateurish technology is layout, where the hapless reader encounters a lack of uniformity in paragraphing, headings, spacing, margins, and other elements of format; authors find it too easy to manipulate parameters and substitute personal whims for established standards, such as those in the *Chicago Manual of Style*. An even worse problem involves the presentation of data in graphs and charts. No program for the preparation of manuscripts has any built-in knowledge of standard formats and important conventions (such as showing the source as part of a table of data); authors invent their own standards. Graphics programs allow bars and regions to be easily decorated with gaudy and extraneous ornamentation, obscuring the information and often presenting it in misleading ways. Edward Tufte's book, *The Visual Display of Quantitative Information*, exposes many of the abuses and prescribes remedies (5). Jon Bentley offers practical advice about document design (6).

A third area under siege is the quality of prose itself. Because authors can generate reasonable facsimiles of professionally published works at home, the old unspoken presumption that a nicely printed work must have undergone a careful quality check by editors is no longer true. There are signs of a growing tension between technical authors and editors; the authors say that the editors are trying to impose arbitrary rules of style, while the editors say that authors think technology automatically makes them good writers.

The computer is slowly but steadily transforming every aspect of publishing. It presents a golden opportunity for more scientists to share their ideas with more of their colleagues; by the same token, it threatens them with an ever-increasing flow of mediocre publications.

### *References*

1. Donald Knuth. 1986. *Computers and Typesetting*. Addison-Wesley. 5 vols. (See especially Vol. A, *The T<sub>E</sub>Xbook*, and Vol. C, *The METAFontbook*.)
2. Wolfgang Horak. 1985. "Office document architecture and office document interchange formats: current status of international standards." *IEEE Computer*. October.
3. Jon Bentley. 1986. "Programming Pearls: Literate Programming." *Com-*

*munications of ACM 29*. May and June.

4. Carolyn Chauncey. 1986. "Special Report: Digital Typography." *Technology Review 89*, 2 (February-March).

5. Edward R. Tufte. 1983. *The Visual Display of Quantitative Information*. Graphics Press.

6. Jon Bentley. 1986. "Document design." *Communications of ACM 29*. September.